Summary of ARC's ORNL DAAC Metadata Findings

April 4th, 2017

This report outlines the ARC team's metadata findings for the ORNL DAAC. The findings reported below are present in the majority of the collection level metadata records. Detailed reports for individual collection and granule level metadata records are provided separately. The detailed reports will be used for tracking the DAAC's progress towards working off the reported findings and metrics will be generated from the detailed reports. An explanation of the color coding found in the detailed reports is included in this report along with preliminary metrics.

1. **Collection Level Findings**
    i. A DOI is currently listed as the short name. The most recent (December 12, 2016) release of UMM-C documentation states that the short name cannot be a DOI. ARC recommends replacing the current DOI value with an appropriate short name. The combination of short name and version number should be unique across CMR.
    ii. Whenever possible, the long name or title should be included in the first sentence of the abstract. Providing the title in the first sentence of the abstract provides clarity for a user who discovers the dataset via the CMR API or has downloaded the metadata directly.
    iii. The 'Collection Progress' element is now a required element in UMM and is meant to describe the production status of the collection. There are three responses for the 'Collection Progress' element and those responses are chosen from a controlled vocabulary list. In order to comply with the controlled vocabulary list, ARC recommends that the "COMPLETED." response be updated to "COMPLETE" in ORNL's collection level metadata.
    iv. Data format information is critical in determining data usability. ARC recommends including data format information whenever possible. When multiple data formats are provided, the data format information should be provided as separate elements in the metadata to support future Earthdata search functionalities.
    v. ARC recommends that the ORNL User Services information be placed in Organization name and that the data center contact information should be removed to consolidate information. An example of using the Organization name is provided here:

<Contact>
<Role>Archiver</Role>
<OrganizationName>DAAC User Services</OrganizationName>
</Contact>

vi. ARC recommends that 'ESIP' and 'EOSDIS' be removed from the 'Project' element. Per UMM documentation, "The project element describes the name of the scientific program, field campaign, or project from which the data were collected. This element may also cover a long term project that continuously creates new data sets - like MEaSUREs from ISCCP and NVAP or CMARES from MISR."

vii. Whenever feasible, the Online Access URL in the metadata should point as directly as possible to the described data. Currently the Online Access URL points to the dataset landing page. Pointing more directly to the appropriate data download folder eliminates confusion for the user and eases data accessibility.

viii. Whenever possible, a link to the dataset landing page should be included for each collection level record. Ideally this link will leverage the DOI URL. More information on DOI landing pages and the information required for a dataset landing page can be found here: https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS/DOI+Landing+Page

ix. ARC recommends adding an online resource URL which links to the ORNL DAAC OGC WMS/WCS service for all relevant collections. Mime type values should be provided for all services and should be selected from the values provided in the UMM-Common documentation. See the detailed report for more information.

x. ARC recommends adding an online resource URL which links to an OPeNDAP resource (THREDDS) for all relevant collections.

xi. ARC recommends providing a description for all online resource URLs. Descriptions should be unique to each URL and not a copy of the dataset abstract. The Browse imagery link would especially benefit from a description that specifies which browse image is provided.

xii. If possible, ARC recommends including the links to the companion files (found on the dataset landing page) in the collection level metadata.

2. Granule Level Findings

i. The granule level data format information does not match the data format of the particular granule. ARC recommends that the data format information match the data format of the specific granule.

ii. Measured parameter information is meant to "expose parameter level quality information about the granule." Therefore, the "Measured Parameters/ Measured Parameter/ Parameter Name" element should only include the name of parameters found directly in the data files. ARC recommends that only parameters found in the data files be included in this element. It should be noted that this field will not be included in the UMM-G in the future. Parameter level metadata will be promoted to its own concept within the CMR via the unified variable model (UMM-Var). See UMM-G documentation for more information.

iii. ARC recommends providing a description for all online resource URLs. Descriptions should be unique to each URL and not a copy of the dataset abstract. For granule level Online Access URL descriptions, ARC recommends removing the abstract and replacing it with a description which indicates to the user that the link provides direct download access of the granule.

iv. Whenever possible, a link to the dataset landing page should be included for each granule level record. Ideally this link will leverage the DOI URL. More information on DOI landing pages and the information required for a dataset landing page can be found here: https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS/DOI+Landing+Page

v. For the OGC WMS/WCS service and the THREDDS data catalog, ARC recommends that a link to the specified granule be provided as an online resource URL in the granule level metadata.

vi. If possible, ARC recommends including the links to the companion files (found on the dataset landing page) in the granule level metadata.

3. Explanation of Color Coding Provided in Detailed Reports

Color	Definition
Cyan	Required field based on UMM-C
Light Purple	An optional primary element with required sub-elements based on UMM-C
Purple	A sub-element which is only required if any information is provided in the scope of the primary element based on UMM-C

White	Completely optional field
Red	Correcting these issues should be of the highest priority
Yellow	Correcting these errors are strongly recommended but are not required
Blue	Minor error/ inconsistency; points out features noticed by the ARC Team which may help improve the robustness of the metadata but are not required to be addressed
*	Any field with an asterisk is controlled by GCMD vocabulary

4. Metrics

15 Collection level records checked

Collection Level	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	263	127	36	930
	28.3%	13.7%	3.8%	

15 Granule level records checked

Granule Level	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	82	32	43	418
	19.6%	7.7%	10.3%	

30 Total records checked (collection + granule)

Cumulative	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	345	159	79	1,348
	25.6%	11.8%	5.9%	